Future Scaling Barriers

Jerry Moench
AMD Fellow
Austin, TX

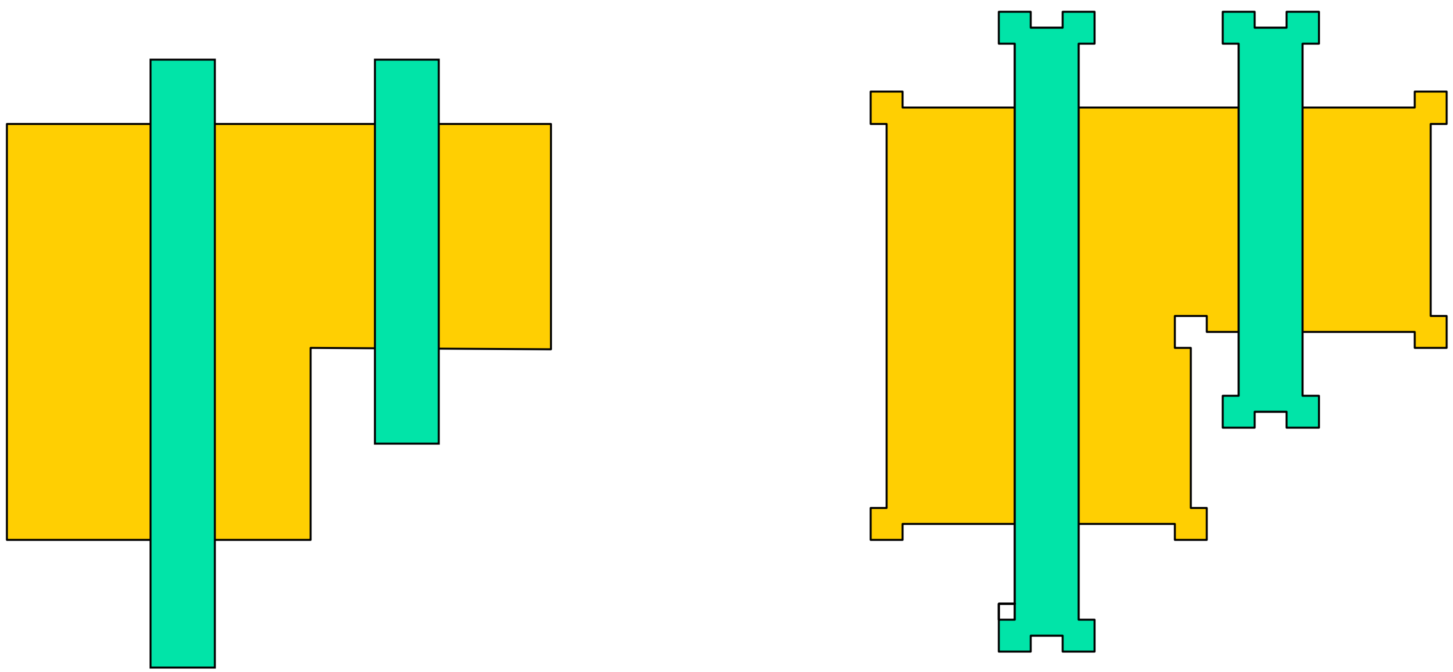
Introduction

- Photolithography limitations
- Transistor leakage
- Gate oxide leakage
- Vt variations
- Metal routing RC limits
- Conclusions

Photolithography limitations

- Currently using 193 nM light
- Resolving ~60 nM shapes
- OPC correction and phase shift masking to allow this to push to the 65 nM technology.
- Next generation stepper 157 nM
- Extreme UV light
- Immersion technology

OPC



Transistor Leakage

- Transistor leakage today dominates standby current
- High Vt on selected transistors
- Wimpy (slightly longer channel)
- Limits circuit design options wide channel connected circuits
- Write ability of dynamic structures and leakage issues could result in no solution

Gate Oxide Leakage

- Every generation gate leakage has increased faster than other leakage
- Gate leakage will become limit scalability of SiO_2
- Gate current on next generation microprocessor could dominate standby current
- Multiple gate oxide thickness and/or new gate oxide material
- Current gate oxide ~4 monolayers

Vt Variations

- Scaling transistor creates more Vt variation
- The number of atoms adjusting Vt 1000's
- Transistor area impacts Vt variation of device
- Vt variation is proportional to 1/sqrt(Area)
- Very small length and width have more variations
- Designs must consider this in many designs
- Monte Carlo simulations are needed
- RAM cell issue on redundancy use

Metal Wiring Improvements

- Improvements in metal have been very slow in developing in the past 30 years.
- Introduction of Cu in Al for better electro-migration.
- Introduction of Cu as a conductor for lower R.
- Low K material for dielectric 2x reduction capacitance.
- Hierarchical metal gives factor of 6-8x improvement.
- Better routing techniques factor of 1.5.
- Total improvement are about ~36x.
- During the same period of time transistor performance has of 1000x improvement!

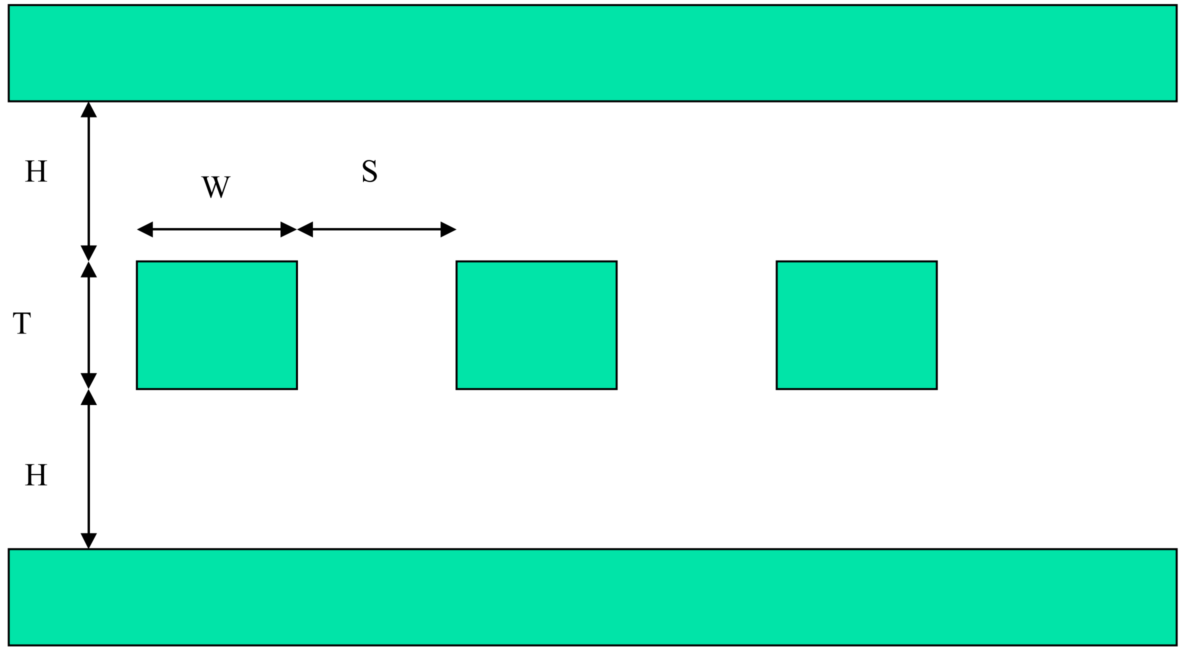
Lower Keff Dielectrics

- The effective K of current 130 nM technology is ~4 Keff best case improvement ~2x to 2 Keff
- Lower K electrical is also lower K thermal.
- Higher temperature will cause higher resistance and more reliability issues due to the temperature.
- Some way to dissipate the power maybe required or at least better modeling of the wire heating is required.

Power, speed power and speed

- Different designs have different problems and thus different solutions.
  - If power dominates then C is most important with reasonable speed. Thus a figure of merit might be Power^2*RC.
  - If speed power product is most important then Power*RC should be optimized.
  - If speed is the only thing important RC needs to be optimized.

Process and design optimization



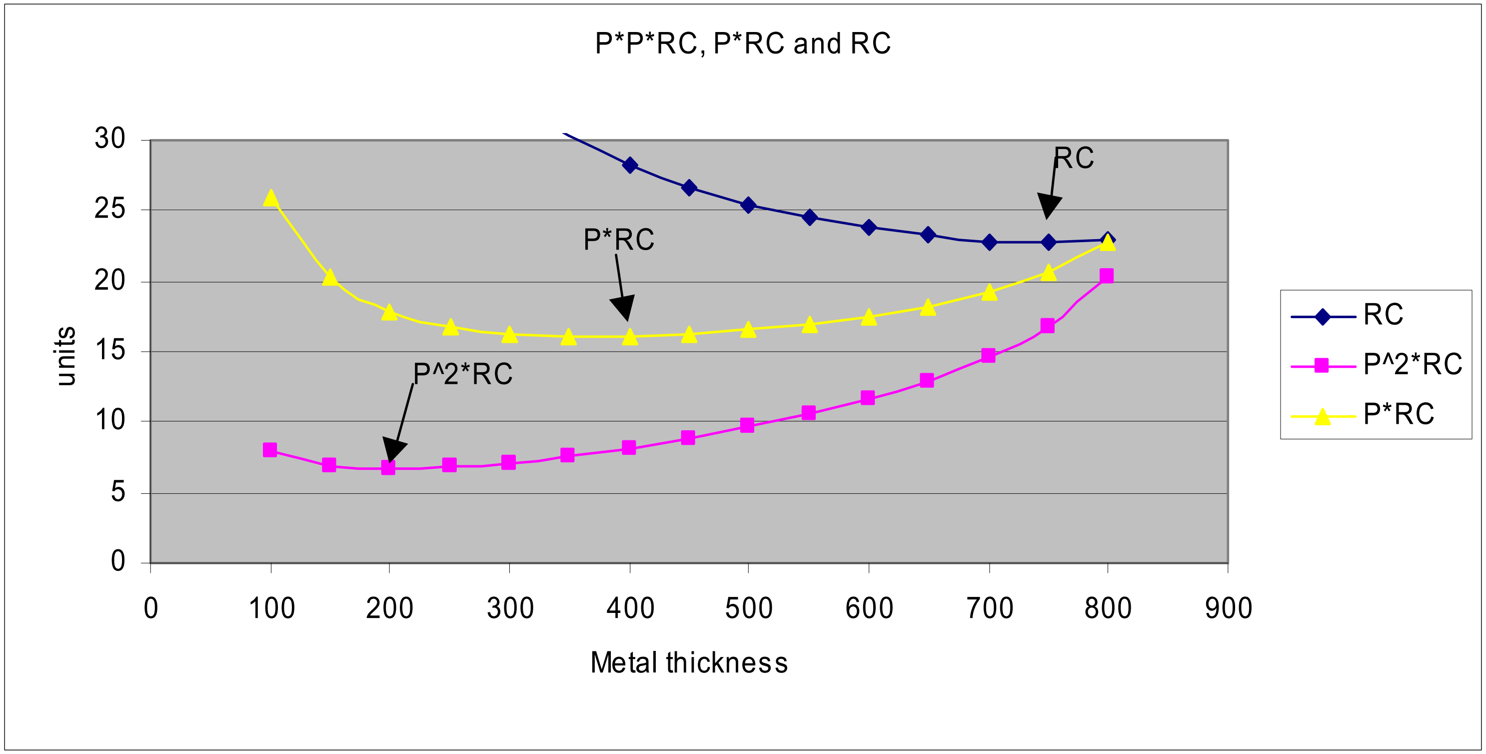
H+T = Constant

H+T = K*(W+S)

Test case for study

- Data for a simple test case will be examined.
- Metal characteristics
  - 1 micron metal pitch 0.5/0.5
  - 1 micron total thickness of metal plus oxide
  - Constant Keff for all dielectrics
  - Constant resistivity for all metal thickness

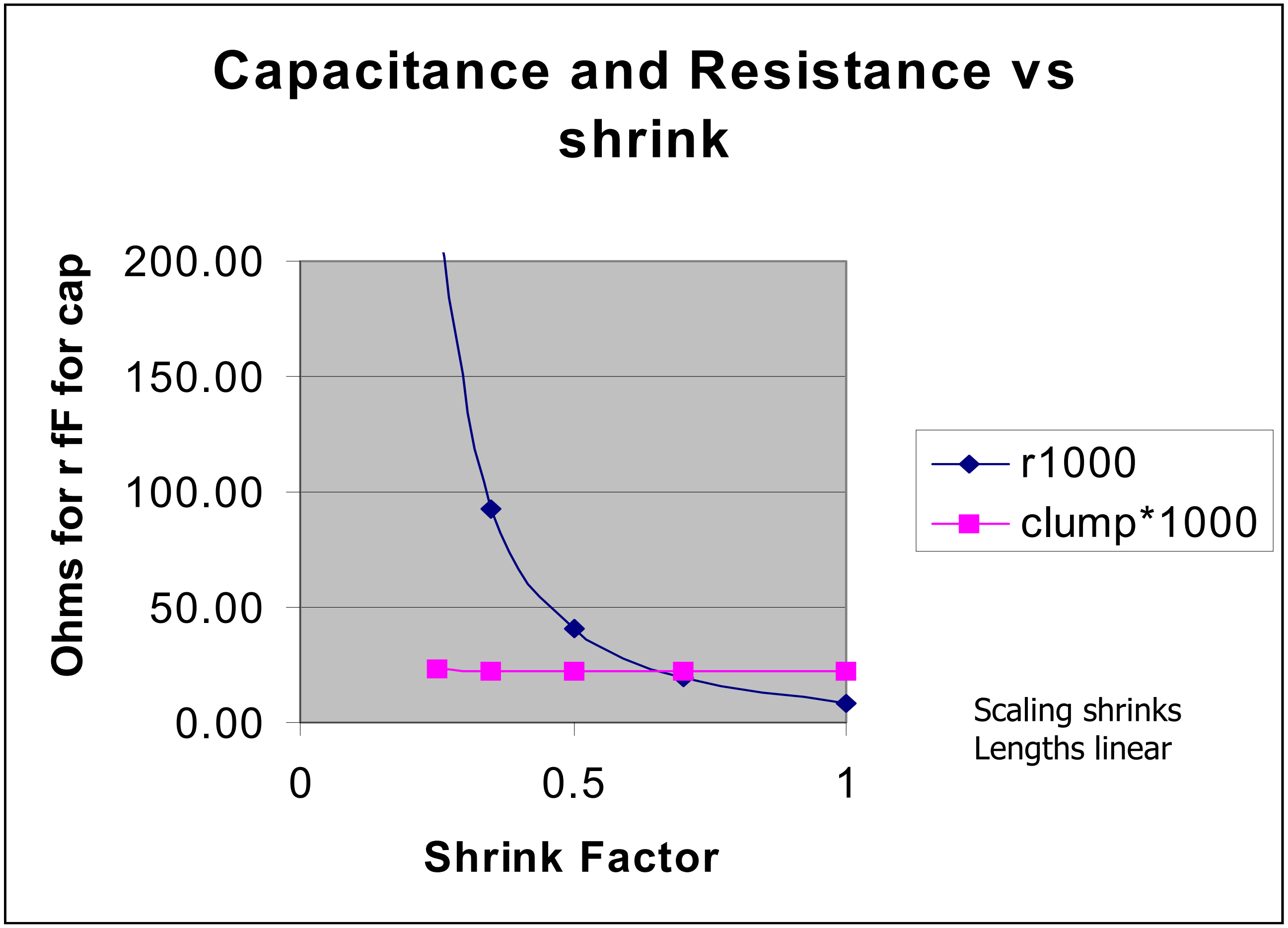
P^2*RC, P*RC and RC



The tradeoff point for high speed design

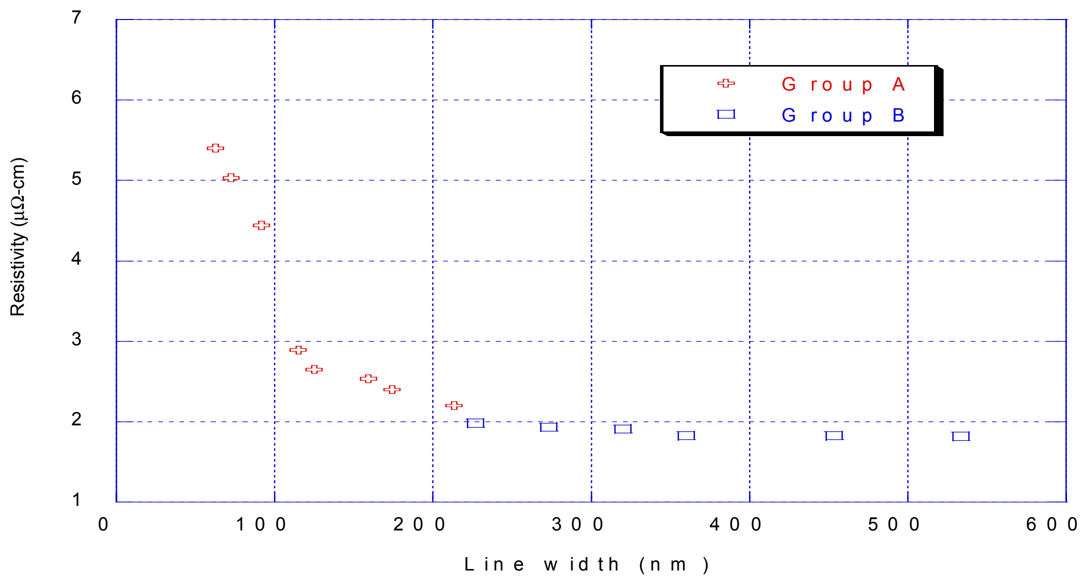
- Balance between best case RC and P*RC
- When the rate of improvement of RC is equal to the rate of increase of P*RC can be considered an optimum compromise point.
- This occurs somewhere between 40% and 75% of the total layer (metal plus oxide) is metal.
- Slight RC improvement comes at a higher rate of power for metal thicker than this point.

Capacitance and Resistance Scaling



SPEED: propagation delay (11)

- Resistivity Vs. Linewidth



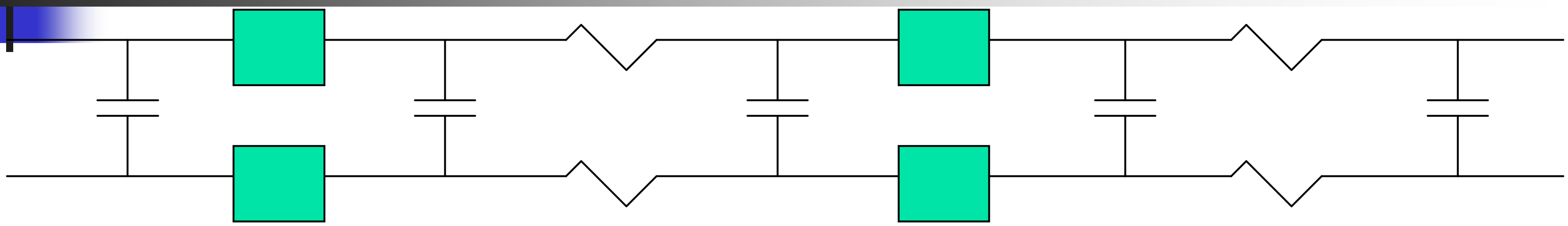
- Inelastic scattering at Cu-barrier sidewall
- Becomes important for dimensions in the order of mean free path of electrons in Cu (10-150nm depending on the constitution of the copper)
- samples with other diffusion barriers show similar trend

Source: W.Wu and K.Maex, IMEC

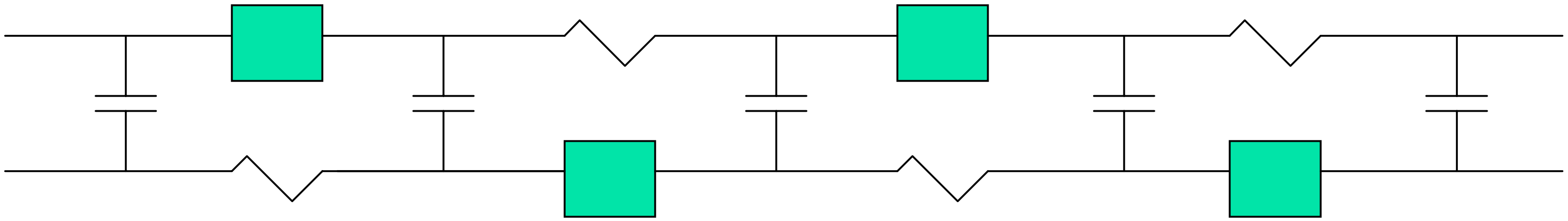
Long line repeater insertion

- On critical static nets other options were used.
  - First approach was to add extra space between lines and/or widen the metal line.
  - Second level was to add extra space and put a shield on one side.
  - Third level was to add extra space/width and put shields on both sides.
- On dynamic long routes shielding and extra spacing was required and signals above and below evaluated for switching.
- Staggered inverter placement was used to reduce noise and aggressor switching.

Long line repeater insertion



Standard Repeater Insertion



Staggered Repeater Insertion

Design checks for coupling issues

- Inductive component increases effective resistance of line plus can contribute to the coupling current spike.
- As edge rates improve the inductive component is a function of ω*L where omega is 2*pi*effective frequency of the edge rate.
- For the design target edge rates in the design, the RLC was typically 10% worse than the simple RC calculation would predict (except on clock signals).
- Assumed both neighbors switching at the fastest edge rate and characterized flops with worst case positioning of pulse.
- For signals going into flops the state node should not be affected by more than 10% due to the input glitch (pulse).
- All nodes were checked for fastest edge rate allowed.

Power Planes

- Power planes can be added between routing layers.
- This gives a good inductive return path for high speed signals.
- It also gives shielding for distribution of clock signals.
- This also gives an improved thermal spreading layer.
- Additional it gives a uniform power distribution network.
- Fairly costly in terms of process complexity.
- Different process problems than normal metal layers.

Routing solutions to reduce noise

- Since there are significant number of power lines a router could be made to route high speed lines next to a power line.
- Additionally long lines should not be run next to each other.
- Stagger repeaters such that true a compliment couple to same signal.
- Insure neighbors are not switching at the same time.
- New router technology to reduce wire length.
- It has been shown that typical routers and placement tools have 100% extra wire because of poor cell placements and routing algorithms.

Routing/Router Issues

- Via resistance and current density through the via are important, if multiple via are required better routers are required.
- Relaxation of metal spaces when possible helps improve RC issues.
- Route lines together that switch at different times in the cycle to improve speed and reduce noise.
- Make every other track long route lines and the others as short route lines when possible.
- Coupling issues will create significant noise problems if not properly addressed.

Architecture changes

- Local modules running at very high speed.
- Synchronization between modules with an extra half clock cycle maybe required.
- Local modules will have clocks gated to reduce power when not active.
- Multiply clock domains running at various frequencies will have synchronization issues.
- Multiple power supplies for various regions so idle modules can be powered down.

Chip Wafer level packaging

- Additional packaging technology at the wafer level could help some of the global issues.
- Other 3-D approaches for stack chips together are being developed and need to be considered .
- Multiple power supplies will be required to power down some regions when not needed.

Conclusions

- Scaling creates a lot of barriers.
- The wiring issues are one of the next major problems on future designs.
- Low K material will present new problems/solutions.
- Each metal layer thickness versus oxide thickness must be optimized to meet the design constraints of speed and power.
- Hierarchical metal will require potential using 10-13 layers of metal on a 45 nM technology.
- Timing driven routers must be improve to optimize routing.
- Architecture of large designs must change to help limit RC problems.
- Wafer level packaging approaches may give additional layers of interconnect.

Backups K7 design data

Design Problem on Athlon Processor

- Metal 2-3 are local interconnect hooking up gates.
- Wire capacitance is the dominate problem and not RC issues.
- Metal 4-5 are longer lines hooking up logic blocks.
- RC and power are more important than the lower metal layers.

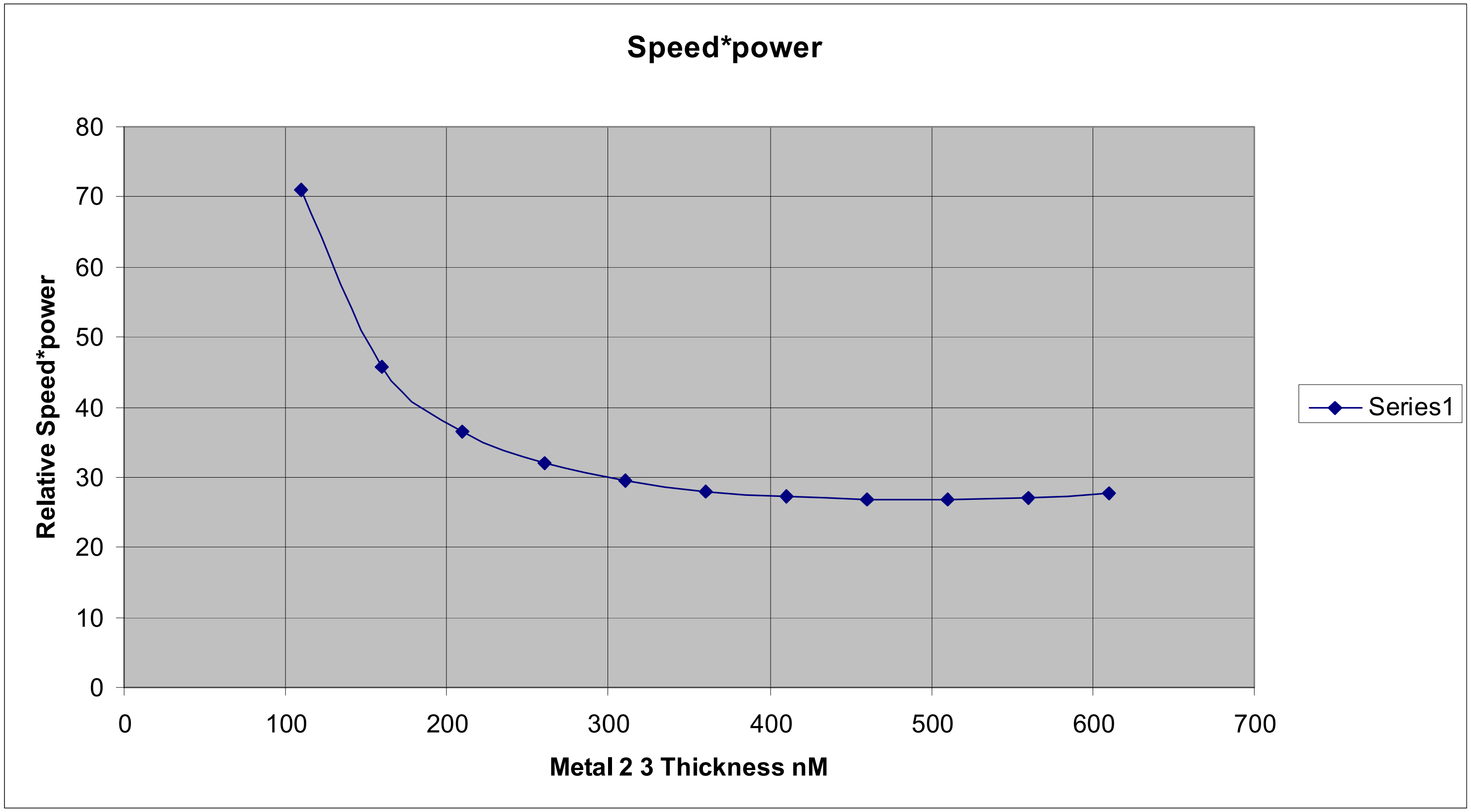
Process and design optimization 130 nM Athlon Processor

- Metal 2-3 drawn width of 0.32 and space of 0.40 microns.
- Nominal thickness (T) of metal 310 nM.
- Nominal oxide thickness (H) 650 nM.
- Metal 4-5 drawn width of 0.45 and space of 0.45 microns.
- Nominal thickness (T) of metal 420 nM.
- Nominal oxide thickness (H) 650 nM.

Metal 2-3 Thickness Optimization

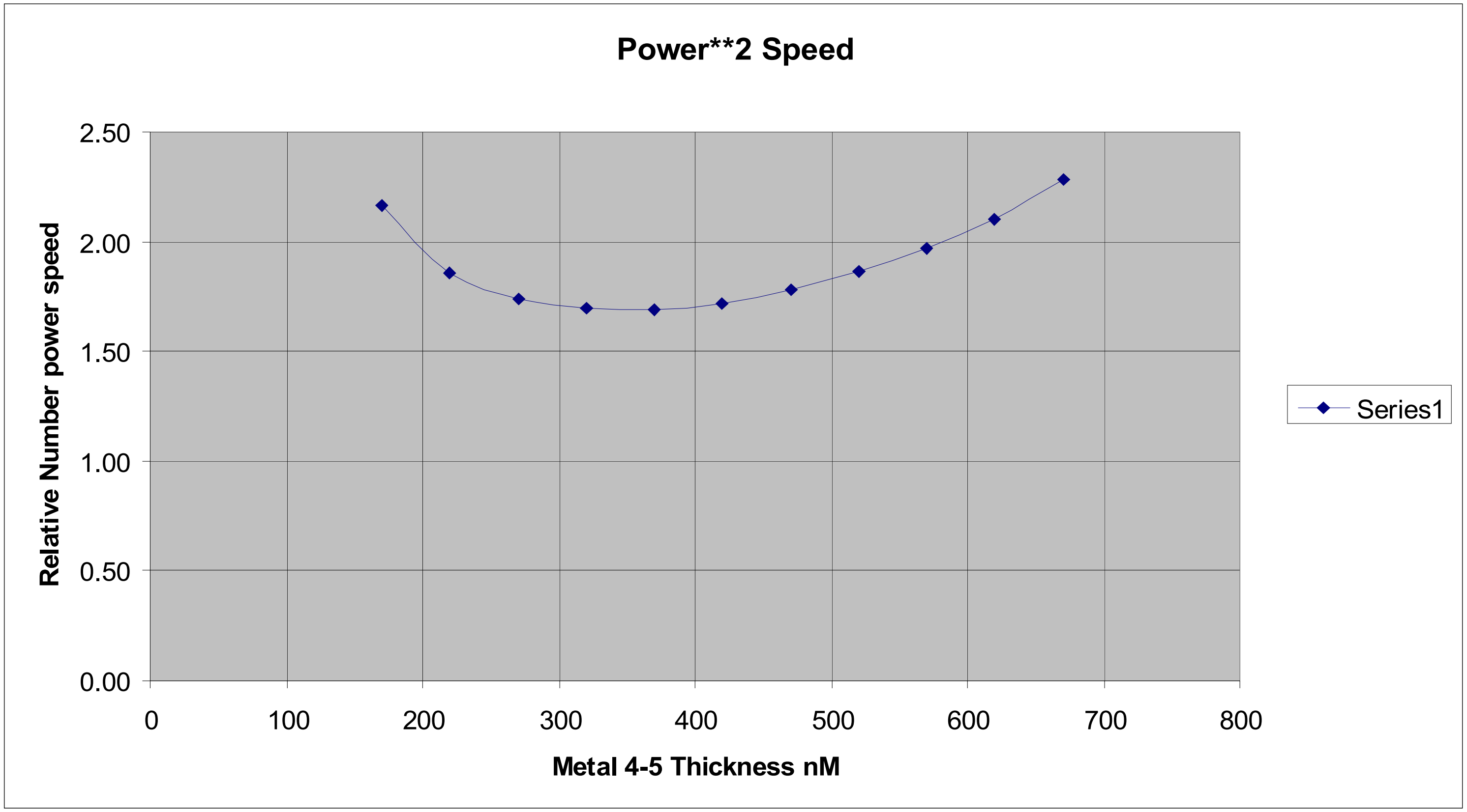
T + H = 960 nM W = 320 nM S = 400 nM

Metal 2-3 Speed Power Vs. Thickness



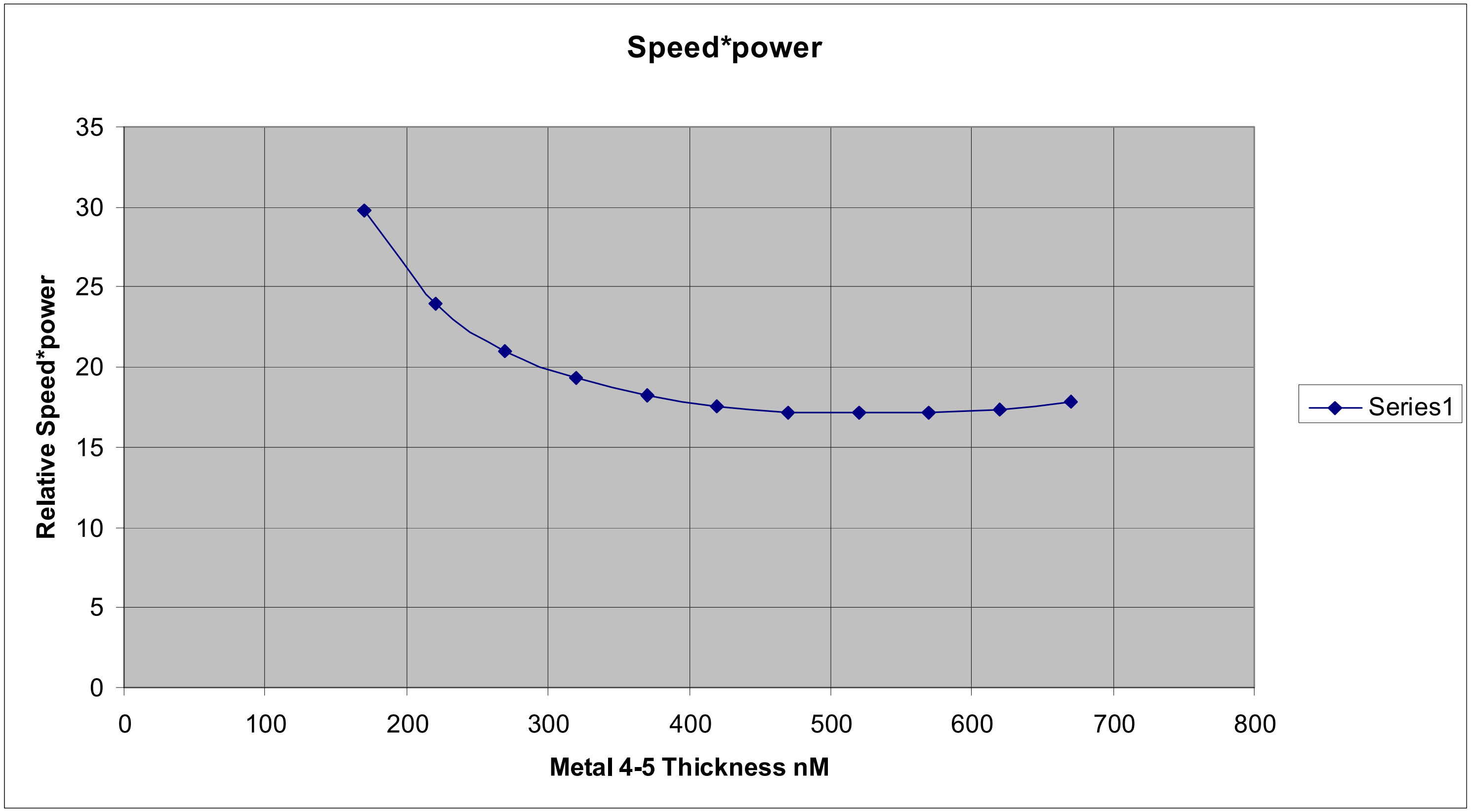
T + H = 960 nM W = 320 nM S = 400 nM

Metal 4-5 power**2 Speed Optimization



T + H = 1070 nM W = 450 nM S = 450 nM

Metal 4-5 Speed Power Vs. Thickness



T + H = 1070 nM W = 450 nM S = 450 nM